CARTA

CARTA: Promoting Computational Research with Art on the Web

Thomas Padilla, Sarah Beth Seymore, Sumitra Duncan, Karl Blumenthal

9/23/2023

From 2021-2023, staff spanning the Internet Archive and the Frick Art Reference Library of the New York Art Resources Consortium (NYARC) worked to foster the development of a national network of art museums and libraries that has come to be known as CARTA (Collaborative ART Archive)[1] under the auspices of the National Endowment for the Humanities supported, *Consortial Action to Preserve Born-Digital, Web-Based Art History & Culture*. Leveraging the nonprofit infrastructure and services of the Internet Archive, 41 diverse art museums and libraries comprising CARTA expanded the ability of art and cultural heritage organizations to collaboratively collect and provide access to essential, ephemeral, born-digital art resources on the web (e.g., art gallery communications, exhibition catalogs, artist websites). To date, CARTA members have preserved and made accessible over 900 web-based art resources, totaling over 13 TB of data with continued growth. In addition to providing unified access to CARTA's collaborative collecting effort through the CARTA Portal,[2] CARTA also promoted the research potential of collections the group developed through a series of workshops (i.e., datathons) and virtual training focused on computational research with web archive data. Given that art libraries as well as a broader community of libraries and museums seek to encourage computational use of their collections, the project team felt it would be useful to share the project team's approach and lessons learned.

Approach

Drawing upon past experience supporting computational research, the Internet Archive and NYARC partner Frick Art Reference Library determined that the educational materials designed to encourage computational research with digital collections should center the following characteristics[3] [4]:

1. **Accessible** - topics introduced in materials are presented at an introductory level with pointers to more advanced resources. No need to design for all possible participant skill levels.
2. **Context-rich** - topics integrate context-rich examples that help motivate participant success in

[1] CARTA https://carta.archive-it.org/
[2] CARTA Portal, https://carta.archive-it.org/collections/.
[3] Internet Archive, Web & Data Services, https://webservices.archive.org/.
[4] https://prattnyarcfellowship.wordpress.com/analysis-evaluation/.

resolving learning objectives and retaining learning outcomes. Real world, discipline specific contextualization aids the instructional effort.

3. **Asset-based** - instructors seek opportunities to highlight the expertise participants bring to the learning experience, rather than assume participants are at a deficit to be addressed by training.[5]
4. **Extensible** - topics and context-rich examples are modular so they can scale from 1 hour to full day engagement. This is in recognition that librarians and/or researchers seeking to promote this form of research need approaches that fit different opportunities.
5. **Sustainable** - materials are well-documented, tools selected are stable, activities are tightly scoped and can be reliably delivered by instructors with varying levels of experience.

In addition to the above characteristics, staff supporting this effort frequently reflected on and aimed to incorporate strategies from *The Carpentries Toolkit of IDEAS* (Inclusion, Diversity, Equity and Accessibility Strategies).[6] Staff were also thankful for and inspired by the datathon model and resources developed by the *Archives Unleashed* project.[7]

Materials developed by project staff are organized into the following modules:

1. **Web archives as data:** introducing the computational research potential of web archives pt. 1
2. **What is a web archive?**: introducing the technical structure of web archives and field of approaches to web archiving generally
3. **Web archiving with Archive-It**: hands on experience conducting web archiving
4. **Evaluating a web archive**: assessing the technical structure of web archives and curatorial decisions that drive their formation
5. **Web archives as primary source**: introducing the computational potential of web archives pt. 2
6. **Web archive data analysis**:
    a. With ARCH (Archives Research Compute Hub)
    b. With Palladio
    c. With Voyant
    d. With python in Google Colab
7. **Coming together** - time for sharing individual and collective lessons learned, next steps for participants post workshop.

---

[5] Heinbach, Chelsea, Brittany Paloma Fiedler, Rosan Mitola, and Emily Pattni. "Dismantling Deficit Thinking: A Strengths-Based Inquiry into the Experiences of Transfer Students in and out of Academic Libraries – In the Library with the Lead Pipe," February 6, 2019. https://www.inthelibrarywiththeleadpipe.org/2019/dismantling-deficit-thinking/.

[6] Robertson, Tara, and Kari L. Jordan. "The Carpentries Toolkit of IDEAS." Zenodo, September 1, 2022. https://doi.org/10.5281/zenodo.7041935.

[7] Milligan, Ian, Samantha Fritz, Nick Ruest, and Jimmy Lin. "Building Community through Archives Unleashed Datathons: Lessons Learned." Presentation. 2021 International Internet Preservation Coalition (IIPC) General Assembly and Web Archiving Conference, June 14-16, 2021, June 15, 2021. https://digital.library.unt.edu/ark:/67531/metadc1827558/m1/2/.

Activities

Project staff held 2 in-person workshops where the above materials were iteratively developed. The venues selected for in-person workshops included the annual National Humanities Alliance Conference (2022) and the annual Art Libraries Society of North America (ARLIS/NA) conference (2023). The rationale for venue selection aligned with an objective to promote the research potential of CARTA collections directly to Humanities researchers (National Humanities Alliance Conference) as well as a community of librarians and museum professionals (ARLIS/NA) who aim to support Humanities researchers doing computational work with CARTA collections.

Learning objectives for the in-person workshops were as follows:

> *During this workshop participants will be (1) introduced to web archives as a primary source, (2) gain familiarity with web archive research use cases; and (3) acquire hands-on experience creating web archive collections and computationally analyzing web archives using Archives Research Compute Hub (ARCH) and a range of other open source tools.*

Each full day, in-person workshop was well-attended, ~25-30 participants respectively. Participants were selected with an eye toward expressed interest in computational research with web archive collections and/or holding a role providing research services in the area. A final criteria for participant selection encouraged diversity of organizational representation in the workshop. A sampling of participant organizations is included below:

Institute of Fine Arts, New York University
George Mason University
American Craft Council
University of San Diego
Shift Collective
National Gallery of Art
Fine Arts Museums of San Francisco
Illinois Wesleyan University
University of New Mexico
University of Georgia
Universidad Nacional Autónoma de México (UNAM)
The Menil Collection
University of Puerto Rico
Ambulante A.C.
National School of Conservation, Restoration and Museography
Wikimedia Foundation
Columbia University Libraries
Getty

Montreal University

In addition to in-person workshops, project staff held dozens of virtual trainings for library, archive, and museum organizations. Internet Archive staff delivered additional in-person variations of workshop materials at the annual International Internet Preservation Consortium (2023) conference and the National Gallery of Art parallel to the Society of American Archivists Conference (2023).

A sampling of participant feedback is included below:

> *As a thought, I think this type of analysis could be for final projects with students or a program for faculty on my campus. I would like to employ some of these methods into my own personal research.*
>
> *It was well organized and provided both theoretical and practical perspectives of web archiving. The presentations and the exercises were very clear so I was able to understand the concepts and background of web archiving and easily explore the tools. As a cataloging and metadata librarian, the use cases from the attendees were also helpful to understand how the archive materials are used in the classroom setting and in art research.*
>
> *In general I was grateful to hear from all three speakers and learned quite a bit from the session. It was very useful to have the morning part of the session to contextualize the tools and methods of analysis we took on in the afternoon. While I was familiar with some of the background content it was important to have it freshly reviewed.*

Use Cases

At a high level, project staff observed four primary use cases common to workshops and virtual training.

Librarian or Museum Professional, Individual, Research Support Objective

A librarian or museum professional working individually, aims to support researchers doing computational work with digital collections. The individual is motivated by current and/or potential researcher needs. The individual aims to support researchers in this space while also recognizing limits on individual capacity given competing commitments. Viability of research service support in this space depends on low-burden, relatively turn-key materials, tools, and services that promote computational use of digital collections.

Librarian, Team, Research Support Objective

A librarian working in concert with a team of other librarians focused on supporting researchers doing computational work with digital collections. The librarian and team are motivated to meet current researcher needs that rationalize their collaboration - e.g., data science, digital scholarship, and data

science initiatives. The librarian and team aim to develop support for web archives research, incorporating that service with related, pre-existing set of services.

Disciplinary Researcher, Pedagogical Objective

A disciplinary researcher is motivated by an opportunity to promote the use of web archives in classroom instruction. The disciplinary researcher aims to promote methods that make use of contemporary primary sources like web archives. Given the need to scale instruction within the undergraduate classroom, the disciplinary researcher relies upon low-burden, relatively turn-key materials, tools, and services.

Disciplinary Researcher, Research Objective

A disciplinary researcher aims to explore research questions with web archives. The disciplinary researcher has well developed research questions yet is faced with the challenge of learning new methods and familiarizing themselves with the nature of web archives as evidence. The disciplinary researcher requires more advanced support as they quickly move beyond introductory topics on web archive research.

Conclusion

As an unparalleled record of contemporary knowledge production, web archives can and will continue to support many lines of research inquiry. Researchers will benefit from coordinated collecting activity by groups like CARTA which results in more representative collection resources. Coordination among organizations in groups like CARTA helps reach a collection scale that would not be possible individually, leveraging complementary resources and expertise. As web archives grow and take on a more central place in disciplinary inquiry it will be important to continue exploring methods and models for promoting computational research with web archives. The Internet Archive, CARTA, and a growing set of partners will continue to refine educational approaches in this space for the benefit of research and education.